

# Economic growth and cancer incidence

1  
3

## 4ABSTRACT

5       Why do we observe increasing rates of new cancer cases? Is the increasing burden of  
6cancer mainly the outcome of higher life expectancy and better life conditions brought about  
7by economic development? To what extent do environmental degradation and changes in life-  
8styles play a relevant role? To answer these questions, we empirically assessed the  
9relationship between per capita income and new cancer cases (incidence) by using cross-  
10sectional data from 122 countries

11       We found that the incidence rate of all-sites cancer increases linearly with per capita  
12income, even after controlling for population ageing, improvement in cancer detection, and  
13omitted spatially correlated variables. If higher incidence rates in developed countries were  
14merely due to those factors, and not also to life-styles and environmental degradation, we  
15would have found a flat or even an inverted-U pattern between per capita income and cancer  
16incidence.

17       The regression analysis was applied also to the eight most common site-specific cancers.  
18This confirmed the existing evidence on the different patterns in rich and poor countries,  
19explained the pattern of the estimated relationship for aggregate cancers, and gave some other  
20interesting insights.

21

22KEYWORDS: Economic development; Cancer; Environmental Kuznets Curve; Environmental  
23degradation; Spatial error models.

25

## 26Highlights

- 27 • New cancer cases increase with p.c. income in a cross-section of 122 countries.
- 28 • Improved detection potential and a longer life alone cannot explain this evidence.
- 29 • Bad life-styles and environmental degradation play a relevant role.

30

31

## 321. Introduction

33Cancer incidence (yearly new cases of cancer) is increasing and predicted to grow fast. The  
34term ‘Cancer epidemic’ has become frequently used, not only by the media (e.g. Servan-  
35Schreiber, 2008), but also by academic journals and by the World Health Organization<sup>1</sup>. The

11<sup>1</sup>In April 2015, the Lancet Oncology and The Lancet launched a joint campaign against cancer “to inform strategies to control  
2the global cancer epidemic” (see <http://www.thelancet.com/campaigns/cancer>). In 2005 the term ‘epidemic’ was used in the 58<sup>th</sup>  
3resolution of the WH assembly, see [http://www.who.int/mediacentre/news/releases/2005/pr\\_wha05/en/](http://www.who.int/mediacentre/news/releases/2005/pr_wha05/en/)

36problem is particularly alarming in lower- and middle-income countries (see, e.g., Boyle and  
37Levin 2008; GLOBOCAN 2012; Stewart and Wild 2014; Vineis and Wild 2014; Ferlay et al.  
382015; Torre et al. 2015). For some rich countries, incidence rates are stabilizing (or slightly  
39decreasing), however at very high rates. In the USA, this has been the case since the mid 1990s  
40(Siegel et al. 2016).

41 Although data availability on cancer has increased significantly in the last years<sup>2</sup>, the  
42relationship between cancer incidence and economic development remains largely  
43unexplored, with just a few exceptions, namely: Beaulieu et al. 2009, Bray et al. 2012, Fidler et  
44al. 2016.<sup>3</sup> The first is a report by “The Economist” Intelligence Unit on the health and economic  
45burden of cancer. As a supplementary result, in one of its appendices, the report shows the  
46outcome of a multiple regression analysis aimed at understanding cross-country variations in  
47both estimated cancer incidence rates for 2009, and in fatality rates for 2002. Regressors  
48included p.c. income, per cent of population aged 65+, and regional dummies. The authors  
49found a positive association of higher cancer incidence rates with both age and higher per  
50capita income countries, which they attributed to the belief of “underreporting of cancer cases  
51in developing countries” (Beaulieu et al., 2009, 62).

52 Bray et al. (2012) and Fidler et al. (2016) grouped countries according to the four levels  
53(low, medium, high, and very high) of the Human Development Index (HDI) and compared  
54incidence and mortality rates across groups. Both articles brought support in favour of the so-  
55called “cancer-transition”, according to which the demographic transition and economic  
56development are changing the composition of the different types of cancers, with a shift from  
57cancers linked to infections to those associated with non-infectious risk factors and possibly  
58associated with the “western” lifestyle.

59 The above-mentioned papers are in line with the health literature, briefly summarised in  
60the next section. The general idea is that increasing cancer incidence rates might be the  
61outcome of economic development, which delivered not only higher life expectancy and  
62improved cancer detection and statistical reporting, but also environmental degradation and  
63“bad” life-styles.

64 The aim of our research was to empirically investigate the macro level relationship  
65between cancer incidence rates and per capita income. For this purpose, we tested some  
66reduced models that looked only at the ends of the complicated causal chains. Such an  
67approach has been followed by the so-called Environmental Kuznets Curve (EKC) literature

---

52 For an assessment of the status of population-based cancer registries worldwide see (Bray et al., 2015).

63 The differences between the present research and the previously mentioned studies will be discussed in section 5.

68that has been investigating the relationship between economic growth and the environment  
69for more than 25 years (e.g., Stern 2004, Dinda 2004, Luzzati 2015). While the EKC literature  
70focused on anthropic pressures, e.g. emissions, here we focused on one possible outcome of  
71pressures, that is, cancer occurrence.

72 The paper is structured as follows. The second section outlines the links between  
73cancer and economic development, from which we derived the conceptual model for our  
74empirical analysis (Figure 1). The third section describes data and methods. In the fourth  
75section results are presented and discussed. The last section gives our conclusions.

76

## 772. Cancer and its possible links with economic development

78 This section firstly summarises what we know about cancer genesis, and then why  
79economic development can play a major role in cancer occurrence. The dominant theory  
80explaining cancer is the so-called Somatic Mutation Theory (SMT) (Nowell 1976; Hanahan and  
81Weinberg 2000 and 2011) according to which “random mutations in the genes which control  
82proliferation or apoptosis are responsible for cancer” (Bertram 2001, p. 170). Hence, cancer is  
83due to stochastic (relevant) mutations that occur in oncogenes and tumour suppressor genes  
84(Lodish et al. 2000). The older a person, the higher is the number of accumulated stochastic  
85mutations, which ultimately leads to higher probability of cancer occurrence.

86 Recently, SMT has been criticised on the basis of theoretical reasons and experimental  
87and epidemiological evidence. Hence, other theories of carcinogenesis have begun to gain  
88ground. They shift the focus from single cells to the entire tissue and attribute a prominent  
89role to altered environments (epigenetic signals) for regulating gene expression, rather than  
90to stochastic mutations of DNA (see e.g. Burgio and Migliore, 2015). For instance, Tissue  
91Organisation Field Theory (TOFT) (see e.g. Baker 2015), which is better seen as integrative  
92rather than alternative to SMT (Bedessem and Ruphy, 2015), looks promising for  
93understanding the role of low-dose foetal exposure to ubiquitous and long lived chemical  
94pollutants, namely the endocrine-disrupting chemicals (EDCs)<sup>1</sup>. These chemicals, by  
95mimicking physiologic hormone signalling molecules, perturbate tightly regulated  
96intercellular signalling pathways. This leads to subtle architectural changes in tissue  
97organization that increase the risk of cancer development. (Howard and Stats 2013).

98 Overall, cancer is increasingly seen as the disruption of a complex equilibrium, that is,  
99the outcome of an evolutionary process in which random genetic mutations have to face the  
100selection of environmental pressure; moreover, intrinsic epigenetic plasticity, clonal evolution

---

81 A useful introduction to EDCs is Gore et al. (2014).

101and high cellular adaptability are also crucial (Greaves, 2014). Hence, cancer is acknowledged  
102as stemming from many interacting factors, that is, from mutations in oncogenes and tumour  
103suppressor genes, from genetic inheritance<sup>2</sup>, work and living environment, and lifestyles (see  
104e.g. Belpomme et al. 2007a, Belpomme et al 2007b, Stewart and Wild 2014).

105 Many studies have investigated the differential contribution to cancer incidence of non-  
106genetic risk factors (e.g. Danaei et al. 2005) and of environmental factors (e.g., Alavanja 2003,  
107Boffetta 2006, Mannucci et al. 2015, Stare and Jozefowicz 2008). The confluence of diverse  
108types of evidence increasingly indicates the relevance of involuntary exposure to  
109environmental contaminants, which affect particularly the “developing foetus, the developing  
110child and adolescent” (Newby and Howard, 2005, 57). For instance, there is evidence of  
111decrease in the average age of cancer onset (e.g. Newby et al. 2007) and increase in childhood  
112cancers (e.g. Steliarova-Foucher et al. 2004), which are also attributed to environmental  
113factors (Stewart and Wild, 2014; Norman et al., 2014). Historical evidence supports the idea  
114that cancer is a disease of industrialization/wealth since “in preindustrial societies, the death  
115rate in infancy was high, but if adolescence was reached then [...] the chances of living a  
116reasonable life span in good health were high and unlikely to end in the development of  
117cancer” (Howard and Statts, 2013). It is not under dispute that economic and technological  
118progress led to the introduction of a complex mixture of persistent xeno-chemicals and other  
119pollutants that have been recognised as carcinogenic.

120 Aggregate quantifications of the environmental risk factors have been proposed in a  
121wide-ranging report by the World Health Organization that surveys the findings on the  
122environmental risk factors (Prüss-Üstün, 2016). According to this report, household and  
123ambient air pollution, passive tobacco smoking, radiation, chemicals, and occupational risks  
124are responsible for at least 20% of cancer cases (in terms of disability-adjusted life years)  
125(Prüss-Üstün, 2016, XVI, 50, and 86).

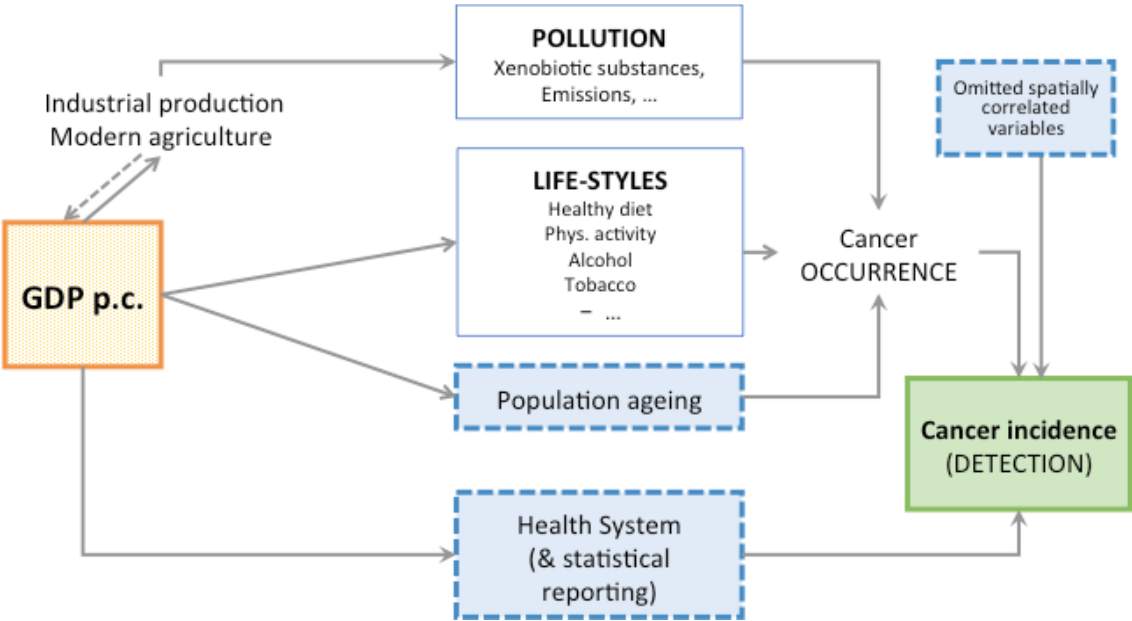
126 Of course, any estimate is highly uncertain (and incomplete) because of the complexity  
127of the cancer-environment relationship in which polluting agents are often time persistent  
128and pervasive, bio-accumulate and are bio-magnificated along the food chain, performing  
129multiple biological actions as well as acting in synergy with other substances. Because of this  
130complexity, we believe that it is also useful to tackle the issue with a very coarse grained  
131perspective, empirically investigating to what extent economic development as a whole plays a  
132role in cancer incidence. To this end, we performed a regression analysis in which income per

---

102 The heritable factors have an important, but not exclusive, role. For instance, using data from Swedish, Danish and Finnish  
11twin registries, it has been reported (Lichtenstein et al 2000) that genetic influence on the incidence of cancer explains no  
12more than 42% of the variance in incidence rate, depending on the cancer site.

capita proxies the joint effects of environmental factors and life-styles on cancer incidence. The reasons why income per capita is expected to be a significant regressor of cancer incidence come from analysing the major links of the causal chain that goes from income to cancer, which are described in what follows and summarised by Figure 1.

137



138

139

140 Figure 1: From income to cancer incidence: major links

141

142 Economic development started with the industrial revolution and was literally fuelled  
143by fossil fuels (e.g., Smil 2000). The availability of an unprecedented quantity of energy  
144radically transformed our economy and the relationship between humans and nature, to the  
145point that many scholars believe that we entered a new epoch, the Anthropocene (Crutzen  
1462002, Steffen et al 2011). A large amount (and number) of pollutants have populated the  
147places where we live, resulting in prolonged and pervasive biochemical stresses that have  
148been found to be risk factors for several diseases, including cancer. Furthermore, other new  
149risk factors (e.g. excessive-weight and obesity) emerged as an outcome of life-style changes  
150that have accompanied the economic development process.

151 At the same time, material living conditions have generally improved, thereby on the one  
152hand bringing about an increase in cancer due to higher life expectancy and, on the other,  
153leading to a reduction in cancers related to some infectious diseases. In other words, income  
154growth has allowed an epidemiological transition<sup>1</sup>, that is, a shift “from a predominance of

141 According to the theory of epidemiological transitions (Omran 2005, pp. 737-738), three ages of mortality patterns in  
15history are observed, namely the age of “pestilence and famine”, of “receding pandemics”, and of “degenerative and man-made  
16diseases”. In the first “age” life expectancy at birth is very low, but epidemic peaks then become less frequent or disappear,

155cancers linked to infections to cancers associated with risk factors that are mainly non-  
156infectious and possibly related to the so-called western lifestyle” (Maule and Merletti 2012, p.  
157745). The identification of this “new epidemiological age” is not only a theoretical construct,  
158but also a relevant empirical fact. According to the World Health Organization (WHO, 2014)  
159about 52% of worldwide deaths in 2012 were due to Non-Communicable Diseases (NCDs)  
160and, of these, about 27% were associated with Malignant Neoplasm.

161 As a concluding remark, it has to be considered that part of the observed increase in  
162cancer incidence along the economic development is the result of improved diagnostic  
163scrutiny and statistical reporting (e.g. Li et al. 2013, Moynihan et al. 2012). In countries where  
164health systems are not well developed, cancer statistics collection is poorly organized and the  
165causes of death often remain undiagnosed resulting in under reporting of cancer deaths in less  
166developed countries (e.g. Fallah and Kharazmi, 2008).

167

### 1683. Material and methods

#### 1693.1 The empirical model

170 The regression model used in this paper is visualised in Figure 1. The items in the  
171dashed contoured boxes have been controlled for in the regressions, so that the variability of  
172incidence rates explained by income can be seen as coming from the joint effect of lifestyles  
173and pollution. The design of the present analysis does not allow a distinction between  
174lifestyles and environmental risks, the importance of which, however, can be drawn from the  
175health literature that was briefly summarised above.

176 Building on the arguments put forward in the previous section, we regressed cancer  
177incidence rates on the 20-year lagged p.c. income while controlling for (1) population ageing,  
178(2) potential for detection and statistical reporting, and (3) omitted factors that might be  
179related to the country’s geographic location.

180 Lags in income were used to take into account the long genesis of cancer and its  
181possible epigenetic nature (see, e.g., Burgio and Migliore, 2015). There are no strong  
182theoretical reasons for taking a particular time lag. For instance, there seems to be a 30-35  
183years lag between the peak in tobacco consumption and the peak in the fatality cases of lung  
184cancer attributable to tobacco smoking (Stewart and Wild 2014, p. 82 ff.; Bilano et al. 2015).  
185This lag is consistent with a lag of 20 years, or more, when considering cancer occurrence. We  
186chose a 20 year lag since it was the longest available time period, due to some lack of older

---

18after which we eventually enter a phase in which mortality tends to approach stability at relatively low levels and non-  
19communicable diseases, including malignant neoplasms, prevail.



187data for income. In any case, we also checked for different time lags (none, 5, 10, and 15  
188years), finding that results do not change qualitatively. This should not be surprising since  
189income is highly autocorrelated. Hence, from an empirical point of view, the choice of the time  
190lag has low relevance. In any case, using lagged values for income is important from a  
191statistical point of view since it avoids potential endogeneity issues.

192 To control for population ageing, we used average standardised rates since they take  
193the different age profiles of countries into account (see below). Furthermore, given that the  
194small size of older age classes in poor countries could cause incidence rates to lose statistical  
195significance, we did a further check by analysing the age class 40-60 separately.

196 Improvements in cancer detection and statistical reporting along the process of  
197economic development were proxied by physician density (physicians per 1000 inhabitants).  
198The reasons for choosing this variable are discussed in detail in the next section.

199 Many other potential factors (such as genetic risk or diet and habits) can be considered  
200as strongly related to the geographic location of the country. Those factors have been omitted  
201since they are either unobservable or lacking in reliable data. A spatial error model was used  
202to take into account these omitted spatially correlated covariates.

### 2033.2 Estimation methods and techniques

204 In order to choose the model that best fits the data, papers within the EKC literature  
205often compare parametric estimates with different specifications of the p.c. income term, i.e.,  
206linear, quadratic or cubic (see Van Alstine and Neumayer, 2010). We followed a different  
207approach. As in Luzzati and Orsini (2009) we preliminarily used semi-parametric methods to  
208assess whether a linear or non-linear specification better fits the data. In the case of evidence  
209of a linear fit, we used it in the parametric estimates. If semiparametric fits suggested non-  
210linear patterns, we chose between the quadratic and cubic specification by minimising the  
211corrected Akaike Information Criterion (AICc) (which also involved maximisation of the  
212adjusted R-squared).

213 For the semiparametric estimates, we used the generalized additive model (GAM), in  
214which each variable enters nonlinearly and separately. We followed the approach proposed in  
215Wood (2006), which is based on penalized regression splines, and used the “mgcv package” in  
216R Development Core Team (2012), with the restricted maximum likelihood (REML) option  
217(see Wood 2011).

218 For the parametric estimates, we followed the spatial econometric methodology  
219developed by Anselin (1988). The reason behind this choice is that, differently from the

inclusion of regional dummy variables (as e.g. in Beaulieu et al. 2009), this methodology explicitly accounts for the effects of spatial correlation due to imperfections in model and measurements that exhibit a spatial structure, thereby increasing the efficiency of the parameter estimates.

The spatial correlation can be incorporated in a regression model in different ways. In the current paper, we used the spatial error model (SEM) in which the spatial correlation is modelled in the error term. This is based on the *a priori* grounds that cultural and genetic factors vary across space and are assumed to drive cancer incidence variability. However, due to lack and/or unreliability of data, such factors are unobserved and/or unmeasured and, therefore, omitted in the regression. If they are influential, then their impact on the explanation of cancer incidence is subsumed in the error term that shows a spatial pattern.

To check if this is the case, a spatial specification search was carried out using OLS estimation and applying a Lagrange Multiplier test. In the presence of evidence of spatial correlation, the spatial model was estimated by means of maximum likelihood.

The SEM was implemented by specifying a spatial stochastic process for the error term, which in turn yields the nonzero correlation for the units that are considered as neighbours. Consequently, the spatial error model requires the definition of a spatial matrix, which reflects the potential interactions between neighbouring units (countries in our case). Here, two different countries are considered as interacting with each other if and only if they belong to the same region. The region taxonomy was taken from the International Agency for Research on Cancer, the specialized cancer agency of the World Health Organization (WHO), and is listed in the appendix. The spatial regressions are estimated using the “spdep package” in R Development Core Team (2012).

### 3.3. Variables

Data on cancer incidence are becoming increasingly reliable due to the diffusion of national cancer registries (see, e.g., Parkin and Donald 2009). However, cross-national differences in coverage and quality of the data collected are quite pronounced, resulting in high variability of both coverage and reliability: thus the quality is often associated with the level of economic development. For a worldwide comparison, the most relevant project is GLOBOCAN, which is today incorporated in “Cancer today”<sup>1</sup>. GLOBOCAN is a project of the International Agency for Research on Cancer of the World Health Organization. Its database contains data for 26 site-specific cancers and for all sites cancer (excluding non-melanoma

---

<sup>221</sup> See <http://globocan.iarc.fr/Default.aspx>



252skin cancer). This project produced the most recent (2012) estimates of incidence, mortality  
253and prevalence.

254 In order to control for differences arising merely from differences in the age profiles of  
255each population, the average standardized rates (weighted) - ASR(W) - have to be used. The  
256standardization procedure (for details see, e.g., Boyle and Parkin 1991) adjusts observed age-  
257specific rates to a reference population, commonly referred to as the Standard Population,  
258usually the world population. The term 'weighted' refers to standard weights taken from the  
259population adopted as a standard. We calculated<sup>2</sup> ASR(W) using the population weight of the  
260World Standard Population<sup>3</sup> and the population data of the United Nations database.

261 The p.c. income variable was the p.c. GDP, expressed in thousands of US\$ PPP2011<sup>4</sup>,  
262taken from the World Bank online database. Income was averaged over three years to mitigate  
263the effect of the business cycle. As stated previously, we used a 20 year lag to consider the long  
264genesis of cancer and tested shorter income time lags, which however left results qualitatively  
265unchanged, as one would expect from the strong autocorrelation of p.c. income (see the  
266appendix, Table A2).

267 As regards the variable to proxy the diagnostic potential of a country, it has to be  
268emphasised that detection and statistical reporting are very different from early detection.  
269While the first two affect the quality of the incidence rates data, the latter is relevant for  
270treating cancer, and hence for mortality rates. Early detection is strongly associated with the  
271presence of screening programmes and diagnostic facilities, which are in turn associated with  
272high levels of per capita health expenditure (and income). For mere detection and statistical  
273reporting, however, an easy access to a doctor is a crucial variable, more important than the  
274availability of advanced technical tools. So far, physician density has proved to be very  
275important in cancer detection (e.g., Ananthakrishnan et al. 2010, Fleisher et al. 2008, Li et al  
2762013, Sundmacher and Busse 2011) and for many other care issues, like infant mortality (e.g.,  
277Farahani et al. 2009), and generally for health outcomes (e.g. Friedberg et al. 2010, Macinko et  
278al. 2007, Mondal and Shitan 2014, Shi 2012). At the same time, physician density can be  
279reasonably thought as having "diminishing returns" in cancer incidence reporting, that is, after

---

242 The database provided by GLOBOCAN already provides ASW(R) rates. Using the data available online and implementing the  
25procedure described by the Glossary section of GLOBOCAN 2012 (<http://globocan.iarc.fr/Pages/glossary.aspx>) we obtained  
26slightly different figures.

273 <http://seer.cancer.gov/stdpopulations/world.who.html> World Standard Population is used also in GLOBOCAN 2012.

284 GDP was taken in Power Purchasing Parity (PPP2011) due to the cross-country nature of the analysis. PPP GDP is gross  
29domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same  
30purchasing power over GDP as a U.S. dollar has in the United States. Data are in current international dollars based on the  
312011 International Comparison Program (ICP).

some thresholds, further increases in the physician density will have increasingly smaller relevance.

For the above reasons, we took physician density as a proxy of cancer incidence reporting potential and used it in the regressions with a concave specification. This was empirically supported by the positive and decreasing marginal impact of physician density on cancer incidence in preliminary semi-parametric estimates (see, e.g., Figure A2). Data for physician density (physicians per 1000 inhabitants) were taken from the World Bank online database. But for a few exceptions, they range from 2010 to 2012. The correlation between physician density and GDP p.c. is not strong enough to prevent the use of both variables as regressors (see Table A2).

We avoided transforming the variables into logarithms since this practice, although common, has been shown to be theoretically weak (Mayumi and Giampietro, 2010). Nonetheless, we verified that using logs does not change the results qualitatively.

#### 3.4. Countries

The GLOBOCAN 2012 dataset covers 184 Countries. We excluded those countries (33) for which data were estimated by merely imputing the data of neighbouring countries or registries in the same area. Of the 151 remaining countries, we excluded 5 that are not included in the World Bank online database from which we took per capita income<sup>1</sup>. We also excluded 18 countries for which 20 year lagged p.c. income or other data were not available. Our final list, presented in the appendix, included 122 countries since six other countries were considered outliers and removed.

As discussed in econometrics textbooks (e.g. Gujarati 2004, 540 ff.), including or excluding outliers is a tricky issue. An outlier differs markedly from the other observations and, hence, “provides a large residual when the chosen model is fitted to the data” (Draper and John 1981, 21). An outlier must be excluded if it is influential, that is, distorting the slope of the regression line or even forcing the researcher to change the model specification. This problem is particularly serious in semiparametric models since “GAMs can be very sensitive to the presence of a small proportion of observations that deviate from the assumed model. In other words: a few atypical observations could seriously affect the non-parametric estimates of the smooth regression function” (Azadeh and Salibian-Barrera, 2011).

A preliminary check on the dataset (see Figure A1) shows that there are some observations that are potentially influential (due to their very high income levels) and for which one can easily imagine that they will have large regression residuals for any

---

<sup>331</sup> State of Palestine, France Guadeloupe, France La Reunion, France Martinique, and France Guyana.

specification that can be conceived. One notices that the observations in question refer to two very small and atypical countries, Singapore (a city-state with a rather idiosyncratic economy) and Luxembourg (whose economy is based on financial services), and to another 5 countries whose economy is strongly based on oil exports<sup>2</sup>. Their special characteristics are such that other countries cannot be thought to mimic their performances. For a formal check of influential observations and outliers we followed the approach developed by Fox (2008)<sup>3</sup>, which is based on studentized residual, hat-values and Cook's distances. Applying it to several model specifications (the linear model and those described in section 4), we found that the 5 "oil" countries and Luxembourg should be excluded, while excluding Singapore is not statistically supported.<sup>4</sup>

323

### 3.5 Data descriptive statistics

A preliminary overview of the data is given by Table A1, which contains the main descriptive statistics for the variables. Table A2 shows the correlation matrix for all-sites cancers, both for the entire population and for the age class 40-60, p.c. income (and its lagged values), and physician density. As expected the autocorrelation of p.c. income is remarkably high.

329

## 4. Results

We start by presenting the results for all cancers, and then we move to organ site-specific cancers. The labels of the variables are as follows. *AllC* refers to incidence rates for all cancers, otherwise the name of site-specific cancer is indicated. The suffix "\_40-60" indicates that the rate refers to the population in the age class 40-60. Incidence rates are measured as yearly new cases on 100,000 inhabitants. *Y\_92* is the three-year average, centred on the year 1992, of GDP p.c. (thousands of \$PPP2011) and *PhysD* is the physician density in 2012 (number of physicians every 1000 inhabitants).

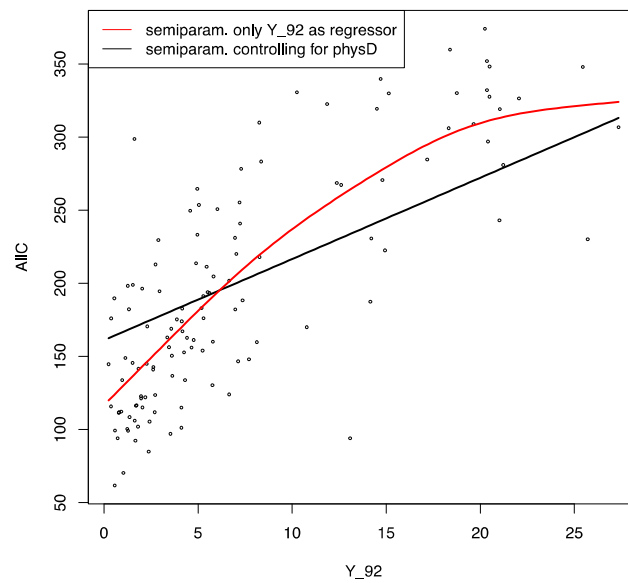
The semiparametric analysis for all cancers is shown in Figures 2a and 2b. Regressing the incidence rate on p.c. income gives the concave curve that is shown in Figure 2a. However, when controlling for physician density, the marginal impact of p.c. income on cancer incidence becomes linear (the straight line in Figure 2a), while the marginal impact of physician density is non-linear (Figure A2). Similar results (not shown) are obtained when analysing the 40-60 age class. Figure 2b shows the confidence bands (5%) of the regression shown in Figure 2a.

---

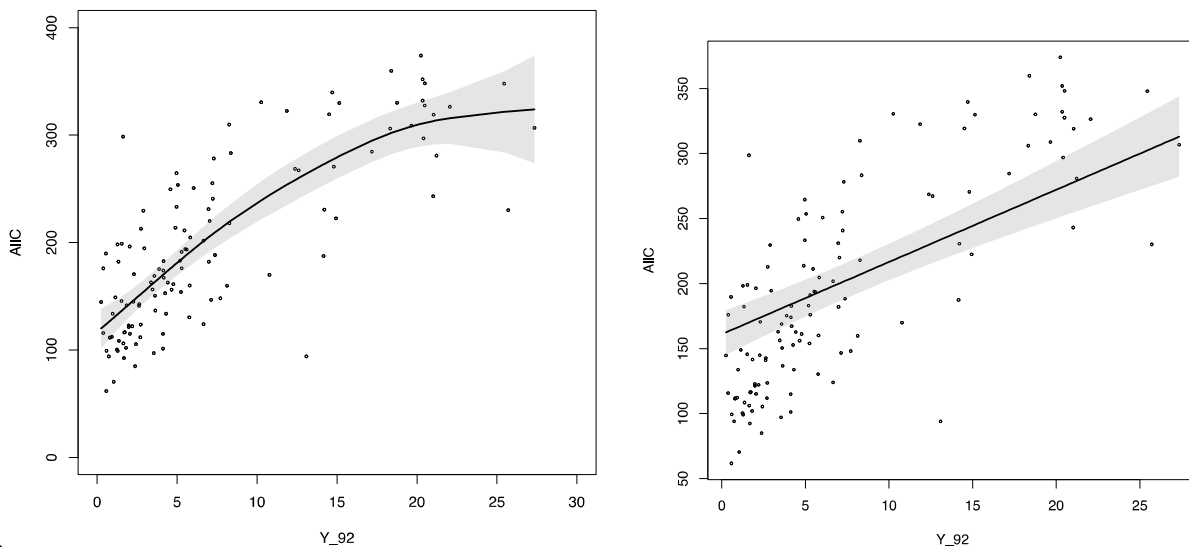
<sup>352</sup> Bahrain, Brunei, Oman, Saudi Arabia, United Arab Emirates.

<sup>363</sup> The approach is also described in Levshina (2015, 153-155). We used the *influencePlot* function in the R-Package 'car' (<https://cran.r-project.org/web/packages/car/car.pdf>).

<sup>384</sup> It is worth noticing that including the 5 "oil" countries requires changing the model specification, while Luxembourg affects only the size of the estimated coefficients.



344  
 345 Figure 2a. Standardised cancer incidence rates in 2012 vs. p.c. income in 1992. Semiparametric fits  
 346 when controlling (straight line) and not controlling (curve) for physician density. Age classes: whole  
 347 population.



348  
 349 Figure 2b. Standardised cancer incidence rates in 2012 vs. p.c. income in 1992. Semiparametric fits  
 350 with confidence bands (5%) when not controlling (left) and controlling (right) for physician density.  
 351 Age classes: whole population.

352 This preliminary evidence, as discussed in detail in section 3.2, was helpful to specify  
 353 the parametric estimates, which, in contrast to the semiparametric estimates, also allow the  
 354 possible spatial correlation of errors to be taken into account.

355 The results of the SEM parametric regressions for all cancers are shown in Table 1. Let  
 356 us start from the first three equations. Equations 1a and 1b refer to the whole population,

357while equation 2 concerns the age class 40-60. In equation 1a income is 20 years lagged, while  
358in eq. 1b income has no lag. The two estimates are very similar. As already mentioned (see  
3593.1), this is due to the high autocorrelation of income, and also holds for estimates referring to  
360other time lags. All estimates show that the incidence rates for all cancers are positively  
361correlated with p.c. income even when controlling for population ageing, physician density  
362and omitted spatially correlated covariates. Standardised incidence rates increase by 4.66 and  
363by 0.64, respectively for all ages (eq. 1a) and for the age class 40-60 (eq. 2), per increase in p.c.  
364income of 1,000\$ (1992, PPP2011). Figure 3 and Figure A3 show the marginal impact on  
365standardised incidence rates of p.c. income, drawn respectively for the whole population (eq.  
3661a) and for the 40-60 age class (eq. 2). It should be noted that p.c. income coefficients may  
367partially capture differences in detection/reporting capacity that cannot be accounted for in  
368terms of physician density. However, as Figure 2 suggests, there are no reasons for believing  
369that including a better proxy would transform the relationship from linear to concave .

370 Table 1. Summary of the Spatial Error Model parametric estimates<sup>1</sup>

<i>Dep. Var.</i>									
<i>AllC</i>	=	108.8	+	4.66		+	47.9	<i>PhysD</i>	- 6.09 <i>PhysD</i> <sup>2</sup> [Eq. 1a]
					<i>Y_92</i>		4		
<i>s.e.</i>		9.71		0.81			8.35		1.41
<i>p.</i>		<0.001		<0.00			<0.00		<0.001
				1			1		
<i>AdjR</i> <sup>2</sup> =0.76; Spatial parameter=0.47, p<0.001									
<i>AllC</i>	=	107.06	+	2.20	<i>Y_12</i>	+	46.06	<i>PhysD</i>	- 5.860 <i>PhysD</i> <sup>2</sup> [Eq. 1b]
<i>s.e.</i>		0.60		0.34			8.06		1.36
<i>p.</i>		<0.001		<0.001			<0.001		<0.001
<i>AdjR</i> <sup>2</sup> =0.78; Spatial parameter=0.49, p<0.01									
<i>AllC_40-60</i>	=	40.05	+	0.64		+	18.5	<i>PhysD</i>	- 2.47 <i>PhysD</i> <sup>2</sup> [Eq. 2]
					<i>Y_92</i>		7		
<i>s.e.</i>		3.30		0.29			2.97		0.50
<i>p.</i>		<0.001		<0.00			<0.00		<0.001
				1			1		
<i>AdjR</i> <sup>2</sup> =0.58; Spatial parameter=0.41, p<0.01									
<i>AllC</i>	=	130.93	+	11.9		-	0.24	<i>Y_92</i> <sup>2</sup>	[Eq. 3]
				4	<i>Y_92</i>				
<i>s.e.</i>		12.04		2.31			0.10		
<i>p.</i>		<0.001		<0.00			0.01		
				1			5		
<i>AdjR</i> <sup>2</sup> =0.72; Spatial parameter=0.58, p<0.01; calculated turning point <i>Y_92</i> =25.402									

371

421 Results are rounded to two decimal places.

Summary		Min	Mean	Max
statistics of the observed values:	<i>AllC</i>	61.8	196.3	374.1
	<i>AllC_40-60</i>	18.3	64.7	115.2
	<i>Y_92</i>	0.251	7.266	27.352
	<i>PhysD</i>	0.02	1.80	6.72

372  
 373 The contribution of physician density to the fitted incidence rates is measured by the  
 374 corresponding terms of Equations 1a and 2 and can be visualised by Figure 4 and Figure A4,  
 375 drawn respectively for the whole population and for the 40-60 age class. For both regressions,  
 376 the impact is positive and increasing only up to roughly 3.8 physicians over 1000 inhabitants.  
 377 Further increases in physician density beyond this value cannot be assessed since very few  
 378 countries surpass it and the confidence bands become very wide. Similar behaviour emerged  
 379 in all the other regressions where physician density was significant.

380

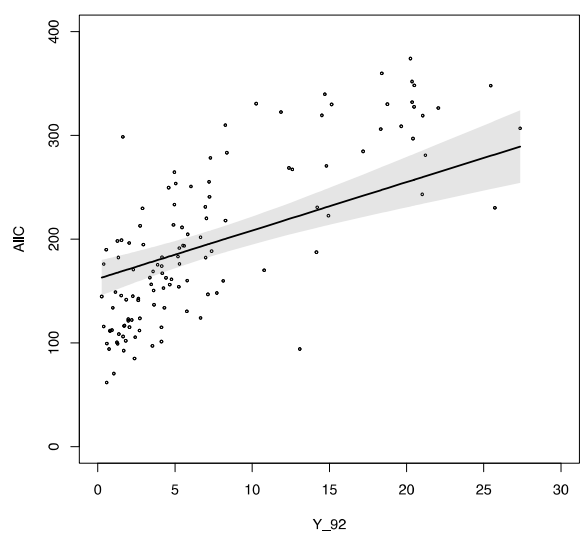


figure 3. Marginal impact (and 95% confidence band)  
 of p.c. income.  
 (Parametric regression of standardized incidence of  
 all-sites cancers on p.c. income and physician  
 density. Age class: all population).

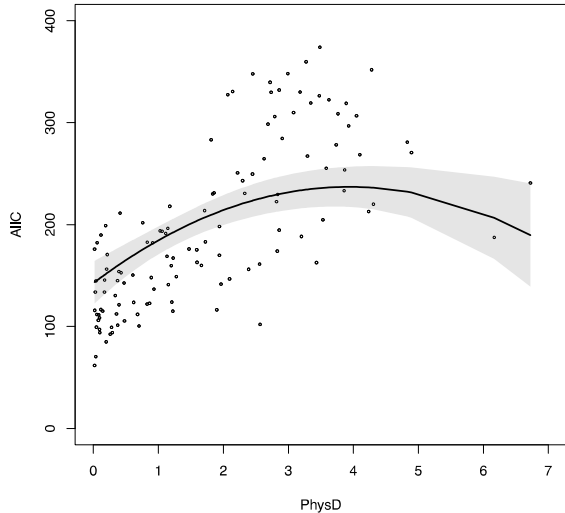


figure 4. Marginal impact (and 95% confidence  
 band) of physician density.  
 (Parametric regression of standardized incidence of  
 all-sites cancers on p.c. income and physician  
 density. Age class: all population).

381  
 382 Equation 3 is shown only to illustrate the effect of not controlling for physician density,  
 383 which however involves mis-specification. Given that the semiparametric fit (see Figure 2)  
 384 was non linear, by minimizing AICc we found that in this case the best specification for  
 385 income was quadratic. The fitted curve has an inverted-U shape. The reason is that the  
 386 quadratic term in equation (3) partially captures that which in equation (1) is captured by  
 387 physician density. The estimated relationship is increasing within almost the entire range of  
 388 the observed income values since the calculated turning point (25,402\$) is close to the highest

389p.c. income (27,352\$) (see Table 1). Not controlling for physician density yields similar  
390consequences in all the other regressions for which the estimated relationship between  
391income and cancer was positive.

392

393 The analysis presented above was replicated for most common site-specific cancers.  
394Table 2 and Table A3 in the appendix, referring respectively to all age classes and to the 40-60  
395age group, summarise the main results. They are organised as follows. Organ site-specific  
396cancers are ordered according to their relative frequency, which is shown in the second  
397column. The third column indicates whether, according to the health literature (see below) the  
398cancer organ site is typical of high-income countries (H) or low/medium income (M-L) ones.  
399Notice that the most frequent organ-site cancers are also typical of high-income countries. The  
400fourth column gives a concise indication of the shape of the estimated relationship while the  
401fifth column indicates whether the estimated income coefficient is positive, negative, non-  
402significant, or follows an EKC shape. The sixth column shows whether the estimated effect of  
403physician density is positive or concave (+), negative (-), or non-significant (n.s). The seventh  
404and eighth columns give the values of the income coefficients and their significance level,  
405while the ninth shows the adjusted R-squared. Full results are in the appendix, Table A4.

406 The prevalence of different organ sites cancer according to the level of development is  
407well established (e.g. Newby et al. 2005; Stewart et al 2014). Lung, breast, colorectum and  
408prostate cancers are the most common organ sites in developed countries, associated both  
409with lifestyles and with environmental factors (Howard and Staats, 2013). Liver, stomach,  
410oesophagus and cervix uteri are highly correlated to chronic infection (such as hepatitis B  
411virus, human papillomaviruses and *Helicobacter pylori*), which are more common in  
412low/medium income countries.

413 When regressing incidence rates on GDP p.c., the sign of the coefficient should be  
414negative for organ-sites typical of low-income countries, and positive (or EKC shaped) for  
415cancers typical of high income countries. This is because increase in GDP in low-income  
416countries improves overall hygiene conditions and thus reduces cancer incidence associated  
417with chronic infection types of cancers. In contrast, in high-income countries an increase in  
418GDP leads to an increase of environmental pollutants and xenobiotic substances and adoption  
419of “western” type life-styles that bring an overall increase in cancer incidence (see discussion  
420in section 2).

421 Our regressions confirmed the expected typicality for developed or LDC countries (see  
422the signs of the income coefficients in Table 2). Results also highlighted that colorectum



cancer is the only one among the most frequent organ-site cancers that follows an EKC shape, although the curve turns down only at rather high levels of p.c. income (at 22,890 1992\$ GDP p.c.). The external risk factors associated with this organ cancer are mainly life-style factors, namely diets that are high in red and processed meats, habitual inactivity, alcohol use, and tobacco smoking. The latter is the most important risk factor for developing lung cancer; however, several environmental and occupational exposures have also been found to be relevant, explaining globally 36% of lung cancer (Prüss-Üstün, 2016, 50). The composition of smoking with other exposures could explain the linearity of the pattern for lung cancer. The environmental risks that are positively associated with income could have offset the beneficial effects of the reduction of cigarette smoking that occurred in many high-income countries (see Stewart et al., 2014).

The colorectum and lung cases highlight the finding that differences in relative frequency between developed and less developed countries (LDCs) do not imply that environmental factors are irrelevant in the organ-site cancers typical of LDCs. For instance, ionizing radiations, exposure in the rubber industry and to asbestos are also risk factors of stomach cancer. Among the sites that we considered, the prostate, the cervix, and the liver are likewise associated with environmental factors (Prüss-Üstün, 2016, 46-51), although the first is typical of high-income countries while the latter two are most frequently found in low-income areas.

In any case, disentangling the environmental effects is difficult since, for each organ site cancer, occurrence is affected by several risk factors that differ according to income. Hence, the actual shape of relationship between income and incidence depends on the relative strength of the various contrasting effects.

Two further remarks can be made. First, physician density was found to be negatively correlated with cervix uteri cancer, which is consistent with the importance of physicians in fostering prevention, and hence reducing incidence rates of this organ site cancer. Second, the estimates for the different organ-site cancers help to understand why a positive relationship with income emerges at the overall level. The reason is that the cancers for which a negative relationship with income holds are less frequent than those for which the relationship is positive.

Table 2. Summary of the Spatial Error Model estimates for the 8 most common organ-site cancers, all age classes

1	2	3	4	5	6	7	8	9
<i>Organ Site</i>	<i>Rel. freq.</i>	<i>Typical of</i>	<i>Model</i>	<i>Role of Y_92</i>	<i>Role of PhysD</i>	<i>Y_92 coeff.</i>	<i>Y_92<sup>2</sup> coeff.</i>	<i>Adj RSq</i>

Lung	13.0%	H	Linear	+	+	0.41**		0.61
Breast	11.9%	H	Linear	+	+	1.89***		0.70
Colorectum	9.7%	H	EKC	\$	+	2.25***	-0.049***	0.76
Prostate	7.9%	H	Linear	+	n.s.	2.43***		0.75
Stomach	6.8%	L/M	Linear	-	+	-0.37***		0.36
Liver	5.6%	L/M	Linear	-	n.s.	-0.36**		0.08
Cervix uteri	3.7%	L/M	Linear	-	-	-0.39^		0.48
Oesophagus	3.2%	L/M	None	none	n.s.	n.s.		0.28

H: High-income countries, L/M: low/medium income countries

Significance levels: \*\*:  $p < 0.05$ , \*\*\*:  $p < 0.01$ , ^: 0.11; n.s.: non-significant

\$ calculated turning point = 22,890 \$ GDP 1992PPP

Finally, the estimated coefficient of the spatial dependence in the error term was positive and significant in all regressions, confirming the relevance of omitted geographically correlated factors.

## Conclusion

The evidence presented in this paper can be compared with the results of three previous studies, already mentioned in the introduction, the primary goal of which, however, was not to explore the relationship between cancer incidence and income growth. Beaulieu et al. (2009) used a methodology similar to ours, that is, they focused on p.c. income and performed a regression analysis. In contrast to us, they controlled for the effect of population ageing by including in the regressions the percentage of population aged 65; additionally, they used intercept dummies to control for geographical differences, but did not control for non-linear influences of income and for cancer reporting improvements. Moreover, they did not use time lags for income. Finally, they had to produce their own estimates of incidence rates, based on GLOBOCAN data for 2002, while we were able to use more reliable and recent data (2012). Hence, their results are not fully comparable with ours. Beaulieu and colleagues interpreted the positive relationship between incidence rates and income as an expected outcome of underreporting cancer cases in developing countries. We show that the positive relationship does not disappear when controlling for physician density, taken as a proxy for incidence reporting. On the contrary, our data and analysis suggest a higher effect of income on cancer incidence rates: in Beaulieu et al. (2009, 63) the income coefficient is equal to 1.457, with a 95% confidence interval of [0.50; 2.4], while ours is equal to 4.66 (see equation 1a), with 95% C.I. of [3.06; 6.26]

Bray et al. (2012) and Fidler et al. (2016) did not use regression analysis or income as a key variable: rather, they compared four groups of countries pooled according to the level of HDI. This could be problematic because of ex-ante defined groups and because HDI also includes life expectancy, which should, instead, be a control variable. In any case, their design

485is too different to allow for a close comparison of the results. Nonetheless, it can be noticed  
486that their papers likewise found both a positive relationship of incidence with levels of  
487development, and different patterns for different cancer sites when comparing less developed  
488and developed countries.

489 Overall, our results are in line with previous evidence, which is not only updated but  
490also strengthened because of the use of a different methodology. Our regressions, which  
491explain a substantial part of the variability (in most cases adjusted R-squared values are  
492higher than 0.6), showed that the relationship between income and cancer incidence rate  
493remains positive (and significant) even after controlling both for favourable effects of  
494economic development - namely population ageing and improvements in cancer detection and  
495statistical reporting - and for spatially correlated omitted variables.

496 Another result of this work is that underreporting can be proxied by a concave function  
497of physician density, the contribution of which is positive and increasing up to roughly 3.8  
498physicians over 1000 inhabitants, while further increases do not seem to be relevant. Only for  
499cervix uter cancer did physician density result to be negatively correlated. This should not be  
500surprising since prevention is particularly important for cervix uter cancer and physicians  
501play a key role in prevention. Finally, omitting to control for physician density would produce  
502in some cases inverted-U patterns, with turning points at the very end (or beyond) of the  
503income range of the sample. Such an omission, however, was regarded here as a  
504misspecification.

505 To sum up, our analysis shows that the cancer epidemic cannot be explained solely by  
506higher life expectancy, by better statistics and by regional peculiarities: rather, a significant  
507role must also be attributed to environmental degradation and life-styles. Unfortunately, our  
508regressions are unable to distinguish between the two. Some clues can be drawn from the case  
509of lung cancer, which, despite the decrease in smokers in high-income countries, still exhibits a  
510positive relationship with income rather than a Kuznets curve. This could be interpreted as  
511arising from environmental exposure. In any case, due to the presence of so many confounding  
512factors, separating the environmental effects from the life-style aspects would require either  
513using micro data or restricting the analysis to units for which a considerable range of statistics  
514is available.

515 However, the relevance of environmental risk factors can be inferred from the  
516increasing evidence available from the health literature according to which “environmental  
517factors play a more important role in cancer genesis than it is usually agreed” (Irigaray 2007,  
518640). Our findings are consistent with this literature, namely, that both social change (e.g.

lifestyles) and “the involuntary exposure to many carcinogens in the environment contributes to the rising trend in cancer incidence” (Belpomme et al. 2007a, 1037).

The policy message that can be drawn from our work is that only by becoming aware of the negative side effects of economic development will we also be able to implement policies to tackle them. This is the message of one of the most important recent reports on cancer, according to which “the realization of just how much disease and ill health can be prevented by focusing on environmental risk factors should add impetus to global efforts to encourage preventive health measures” (Prüss-Üstün, 2016, VII).

527

## 528ACKNOWLEDGMENTS

We gratefully acknowledge comments and suggestions on earlier versions of this paper received from A. Di Ciaula, D. Fiaschi, C. V. Howard, F. Merletti and colleagues, G. Nebbia, and from participants in the International Society of Ecological Economics (ISEE) 2016 Conference (University of the District of Columbia, Washington DC, June 26-29 2016) and in the “Economics, Economic Policies and Sustainable Growth in the Wake of the Crisis” (Università Politecnica delle Marche, Ancona, Italy, September 8-10 2016). We truly appreciated the work done by the anonymous reviewers who have helped improve the original draft considerably. Errors remain our own.

The present research was partially funded by a PRA 2015 and PRA 2017 grant of the University of Pisa. The paper is the outcome of the joint work of the authors. Nonetheless, sections 1, 3.1, 3.2, and 5 are attributable to Luzzati, sections 3.2, 3.4 to Parenti, sections 2 and 3.5 to Rughi. The latter also prepared the dataset (including standardisation) and performed a preliminary OLS analysis.

541

## 542References

543

Alavanja MC, Samanic C, Dosemeci M, Lubin J, Tarone R, Lynch CF, Knott C, Thomas K, Hoppin JA, Barker J, Coble J, Sandler DP, Blair A, (2003). Use of agricultural pesticides and prostate cancer risk in the Agricultural Health Study cohort. *American Journal of Epidemiology*, 157(9), 800-814.

Ananthakrishnan AN, Hoffmann RG, & Saeian K (2010). Higher physician density is associated with lower incidence of late-stage colorectal cancer. *Journal of general internal medicine*, 25(11), 1164-1171.

Anselin L (1988). *Spatial Econometrics: Methods and Models*. Kluwer: Dordrecht.

Azadeh A and Salibian-Barrera M. (2011). An outlier-robust fit for generalized additive models with applications to disease outbreak detection. *Journal of the American Statistical Association* 106.494, 719-731.

Baker SG. A cancer theory kerfuffle can lead to new lines of research (2015). *Journal of the National Cancer Institute*, 107(2), 1-8.

Beaulieu N, Bloom D, Bloom R, and Stein R (2009). Breakaway: The global burden of cancer—challenges and opportunities. The Economist Intelligence Unit.

559 Bedessem B, Ruphy S. (2015) SMT or TOFT? How the two main theories of carcinogenesis are  
560 made (artificially) incompatible. *Acta Biotheor.* 63(3):257-67.,

561 Belpomme D, Irigaray P, Hardell L, Clapp R, Montagnier L, Epstein S, & Sasco AJ (2007a). The  
562 multitude and diversity of environmental carcinogens. *Environmental research*, 105(3),  
563 414-429.

564 Belpomme D, Irigaray P, Sasco AJ, Newby JA, Howard V, Clapp R, Hardell L (2007b). The  
565 growing incidence of cancer: role of lifestyle and screening detection, *Int J Oncol*, 30,  
566 1037-1049

567 Bertram JS (2001). The molecular biology of cancer. *Molecular aspects of medicine*, 21(6),  
568 167-223.

569 Bilano V, Gilmour S, Moffiet T, d'Espaignet ET, Stevens GA, Commar A, Tuyl F, Hudson I, and  
570 Shibuya K, 2015. Global trends and projections for tobacco use, 1990–2025: an analysis  
571 of smoking indicators from the WHO Comprehensive Information Systems for Tobacco  
572 Control. *The Lancet*, 385 (9972), pp.966-976.

573 Boffetta P (2006). "Human cancer from environmental pollutants: the epidemiological  
574 evidence." *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 608.2  
575 (2006): 157-162.

576 Boyle P, Levin B (2008). *World Cancer Report 2008*. World Health Organisation.

577 Boyle P, and Parkin DM, (1991) "Statistical methods for registries." *Cancer registration:*  
578 *principles and methods* 95, 126-158.

579 Bray F, Ferlay J, Laversanne M, Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M,  
580 Steliarova-Foucher E, Swaminathan R, Antoni S and Soerjomataram I (2015). Cancer  
581 incidence in five continents: inclusion criteria, highlights from volume X and the global  
582 status of cancer registration. *International Journal of Cancer*, 137(9), pp.2060-2071.

583 Bray F, Jemal A, Grey N, Ferlay J, and Forman D (2012). Global cancer transitions according to  
584 the Human Development Index (2008–2030): a population-based study. *The lancet*  
585 *oncology*, 13(8), 790-801.

586 Burgio E, Migliore L, 2015. Towards a systemic paradigm in carcinogenesis: linking epigenetics  
587 and genetics. *Molecular biology reports* 42.4 pp. 777-790.

588 Crutzen, PJ *Geology of mankind*. *Nature* 415, no. 6867 (2002), pp. 23-23.

589 Danaei G, Vander Hoorn S, Lopez AD, Murray CJ, Ezzati M, & Comparative Risk Assessment  
590 collaborating group (Cancers) (2005). Causes of cancer in the world: comparative risk  
591 assessment of nine behavioural and environmental risk factors. *The Lancet*, 366(9499),  
592 1784-1793.

593 Dinda S (2004). Environmental Kuznets curve hypothesis: a survey. *Ecological economics*,  
594 49(4), 431-455.

595 Draper NR, John JA. Influential observations and outliers in regression. *Technometrics*. 1981  
596 Feb 1;23(1):21-6.

597 Fallah M and Kharazmi E (2008). Substantial under-estimation in cancer incidence estimates  
598 for developing countries due to under-ascertainment in elderly cancer cases. *Cancer*  
599 *Letters* 2008; 264: 250–255.

600 Farahani M, Subramanian SV, and Canning D (2009). The effect of changes in health sector  
601 resources on infant mortality in the short-run and the long-run: a longitudinal  
602 econometric analysis. *Social Science & Medicine*, 68(11), 1918-1925.

603 Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, and  
604 Bray F (2015). Cancer incidence and mortality worldwide: Sources, methods and major  
605 patterns in GLOBOCAN 2012. *Int. J. Cancer*, 136: E359–E386. doi:10.1002/ijc.29210

606 Fidler MM, Soerjomataram I, and Bray F (2016). A global view on cancer incidence and  
607 national levels of the human development index. *Int. J. Cancer*, 139: 2436–2446.

608 Fleisher JM, Lou JQ, and Farrell M, (2008). Relationship between physician supply and breast  
609 cancer survival: a geographic approach. *Journal of community health*, 33(4), 179-182.

610 Friedberg MW, Hussey PS, and Schneider EC (2010). Primary care: a critical review of the  
611 evidence on quality and costs of health care. *Health Affairs*, 29(5), 766-772.

612 GLOBOCAN 2012. Estimated cancer incidence, mortality and prevalence worldwide in 2012,  
613 WHO, <http://globocan.iarc.fr/Default.aspx>

614 Gore AC, Crews D, Doan LL, La Merrill M, Patisaul H, Zota A (2014). Introduction to endocrine  
615 disrupting chemicals (EDCs), Endocrine Society and Ipen, available on line at  
616 [https://saludsindanio.org/sites/default/files/documents-files/2999/Introduction](https://saludsindanio.org/sites/default/files/documents-files/2999/Introduction%20to%20Endocrine%20Disrupting%20Chemicals.pdf)  
617 [%20to%20Endocrine%20Disrupting%20Chemicals.pdf](https://saludsindanio.org/sites/default/files/documents-files/2999/Introduction%20to%20Endocrine%20Disrupting%20Chemicals.pdf)

618 Greaves M, 2014. An evolutionary foundation for cancer control” in Stewart B and Wild CP  
619 (2014)

620 Gujarati, DN, 2004. Basic econometrics. The McGraw– Hill.

621 Hanahan D, Weinberg RA (2000). "The hallmarks of cancer", *Cell* 7, 100: 57-70.

622 Hanahan D, Weinberg RA (2011). Hallmarks of cancer: the next generation. *Cell* 4(144):646–  
623 674.

624 Howard CV, and Staats G (2013). Cancer and Environmental Influences. *Encyclopedia of*  
625 *Systems Biology*, 172-179.

626 Irigaray P, Newby JA, Clapp R, Hardell L, Howard V, Montagnier L, Epstein S, Belpomme D.  
627 (2007) Lifestyle-related factors and environmental agents causing cancer: an overview.  
628 *Biomedicine & Pharmacotherapy*. 61(10):640-58.

629 LeSage J, and Pace RK (2009). *Introduction to Spatial Econometrics*. Boca Raton, Florida: CRC  
630 Press, Taylor & Francis Group.

631 Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*.  
632 John Benjamins Publishing Company.

633 Li N, Du XL, Reitzel LR, Xu L and Sturgis EM (2013). Impact of enhanced detection on the  
634 increase in thyroid cancer incidence in the United States: review of incidence trends by  
635 socioeconomic status within the surveillance, epidemiology, and end results registry,  
636 1980–2008. *Thyroid*, 23(1), pp.103-110.

637 Li N, Du XL, Reitzel LR, Xu L and Sturgis EM (2013). Impact of enhanced detection on the  
638 increase in thyroid cancer incidence in the United States: review of incidence trends by  
639 socioeconomic status within the surveillance, epidemiology, and end results registry,  
640 1980–2008. *Thyroid*, 23(1), pp.103-110.

641 Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J., Koskenvuo M, Pukkala E, Skytthe A,  
642 Hemminki K (2000). Environmental and heritable factors in the causation of cancer—  
643 analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal*  
644 *of Medicine*, 343(2), 78-85.

645Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, & Darnell J (2000). *Molecular Cell*  
646 *Biology* WH Freeman. *New York*, available at  
647 <http://www.ncbi.nlm.nih.gov/books/NBK21662/>

648Luzzati T, and Orsini M (2009). "Investigating the energy-environmental Kuznets curve."  
649 *Energy* 34.3: 291-300.

650Luzzati T (2015). Kuznets Curves, in Wright J.D., *International Encyclopedia of the Social &*  
651 *Behavioral Sciences*, 2nd edition, Vol 13, 144-149, Elsevier.

652Macinko J, Starfield B, and Shi L (2007). Quantifying the health benefits of primary care  
653 physician supply in the United States. *International Journal of Health Services*, 37(1),  
654 111-126.

655Mannucci PM, Harari S, Martinelli I, and Franchini M (2015). Effects on health of air pollution:  
656 a narrative review. *Internal and Emergency Medicine*, 10(6), 657-662.

657Maule M, and Merletti F (2012). Cancer transition and priorities for cancer control. *The Lancet*  
658 *Oncology*, 13(8), 745-746.

659Mayumi K, and Giampietro M (2010). Dimensions and logarithmic function in economics: A  
660 short critical analysis. *Ecological Economics*, 69(8), 1604-1609.

661Mondal MNI, and Shitan M (2014). Relative importance of demographic, socioeconomic and  
662 health factors on life expectancy in low-and lower-middle-income countries. *Journal of*  
663 *Epidemiology*, 24(2), 117-124.

664Moynihan R, Doust J, and Henry D (2012). Preventing overdiagnosis: how to stop harming the  
665 healthy. *Bmj*, (e3502).

666Newby JA and Vyvyan Howard C. (2005). Environmental influences in cancer aetiology. *Journal*  
667 *of Nutritional & Environmental Medicine*. 5(2-3), 56-114.

668Newby JA, Busby CC, Howard CV, Platt MJ. (2007), The cancer incidence temporality index: an  
669 index to show temporal changes in the age of onset of overall and specific cancer  
670 (England and Wales, 1971–1999). *Biomedicine & Pharmacotherapy*, 61(10), 623-30.

671Norman RE, Ryan A, Grant K, Sitas F, and Scott JG (2014). Environmental contributors to  
672 childhood cancers. *J Environ Immunology and Toxicology*, 1(4), 12.

673Nowell PC (1976). The clonal evolution of tumor cell populations. *Science*, 194(4260), 23-28.

674Omran AR (2005). The epidemiologic transition: a theory of the epidemiology of population  
675 change. *Milbank Quarterly* 83.4: 731-757.

676Parkin DM (2006). The evolution of the population-based cancer registry." *Nature Reviews*  
677 *Cancer* 6.8: 603-612.

678Prüss-Üstün A, Wolf J, Cor Prüss alán C, Bos R, Neira M (2016). Preventing disease through  
679 healthy environments. World Health Organization, Geneva, Switzerland.

680R Development Core Team. (2012). *R: A Language and Environment for Statistical Computing*.  
681 *R Foundation for Statistical Computing*, Vienna, Austria. URL <http://www.R-project.org/>.  
682 ISBN 3-900051-07-0.

683Servan-Schreiber D (2008). We can stop the cancer epidemic, *The New York Times*, Sept. 19,  
684 2008, available at [http://www.nytimes.com/2008/09/19/opinion/19iht-](http://www.nytimes.com/2008/09/19/opinion/19iht-edservan.1.16308287.html)  
685 [edservan.1.16308287.html](http://www.nytimes.com/2008/09/19/opinion/19iht-edservan.1.16308287.html)



686Stare SM, & Jozefowicz JJ (2008). The effects of environmental factors on cancer prevalence  
687 rates and specific cancer mortality rates in a sample of OECD developed countries.  
688 International Journal of Applied Economics, 5(2), 24.

689Shi L (2012). The impact of primary care: a focused review. Scientifica, 2012.

690Siegel RL, Miller KD, & Jemal A (2016). Cancer statistics, 2016. CA: a cancer journal for  
691 clinicians, 66(1), 7-30.

692Smil V (2000). Energy in the twentieth century: resources, conversions, costs, uses, and  
693 consequences. Annual Review of Energy and the Environment, 25(1), 21-51.

694Steffen W, Grinevald J, Crutzen P, and McNeill J (2011). The Anthropocene: conceptual and  
695 historical perspectives. Philosophical Transactions of the Royal Society of London A:  
696 Mathematical, Physical and Engineering Sciences, 369(1938), 842-867.

697Steliarova-Foucher E, Stiller C, Kaatsch P, Berrino F, Coebergh JW, Lacour B, and Perkin M  
698 (2004). Geographical patterns and time trends of cancer incidence and survival among  
699 children and adolescents in Europe since the 1970s (the ACCIS project): an  
700 epidemiological study. The Lancet, 364(9451), 2097-2105.

701Stern DI (2004). The rise and fall of the environmental Kuznets curve. World development,  
702 32(8), 1419-1439.

703Stewart B and Wild CP (eds.) (2014). World cancer report 2014. World Health Organisation,  
704 Geneva Switzerland.

705Sundmacher L and Busse R (2011). The impact of physician supply on avoidable cancer deaths  
706 in Germany. A spatial analysis. Health Policy, 103(1), 53-62.

707Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, and Jemal A (2015). Global cancer  
708 statistics, 2012. CA: A Cancer Journal for Clinicians, 65: 87–108.

709Van Alstine J and Neumayer E (2010) “The environmental Kuznets curve”. In: Gallagher KP  
710 (ed.) Handbook on trade and the environment, E. Elgar.

711Vineis P, and Wild CP (2014). Global cancer patterns: causes and prevention. The Lancet,  
712 383(9916), 549-557.

713WHO (2014). Global status report on noncommunicable diseases 2014. World Health  
714 Organization, Geneva, Switzerland.

715Wood SN (2006). Generalized Additive Models. An Introduction with R. Chapman and Hall,  
716 London.

717Wood SN (2011). Fast stable restricted maximum likelihood and marginal likelihood  
718 estimation of semiparametric generalized linear models. Journal of the Royal Statistical  
719 Society (B), 1(73), 3–36.

720

721

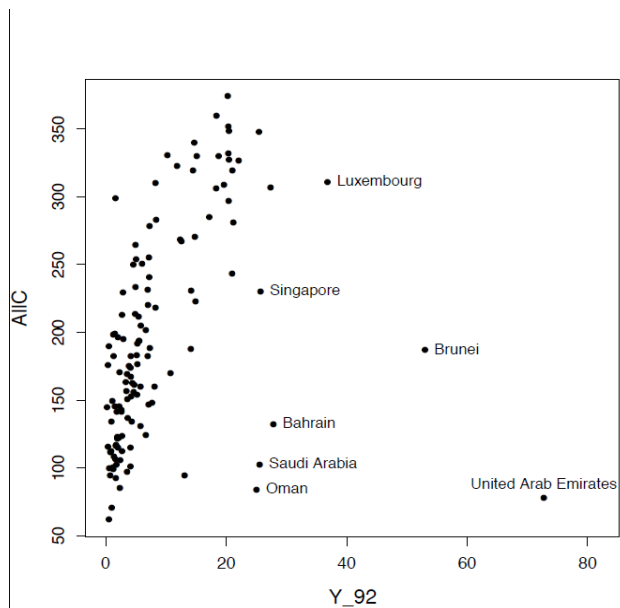


Figure A1. Scatter plot of standardized incidence of all-sites cancers vs. p.c. income in 1992: outliers.

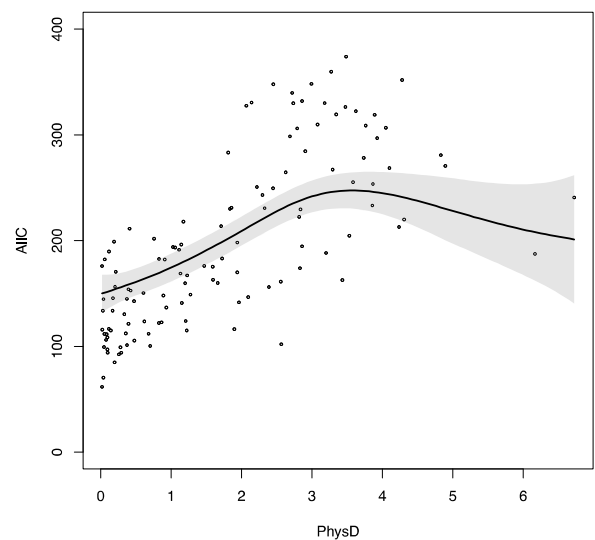


Figure A2. Marginal impact of Physician density in the semi-parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: all population.

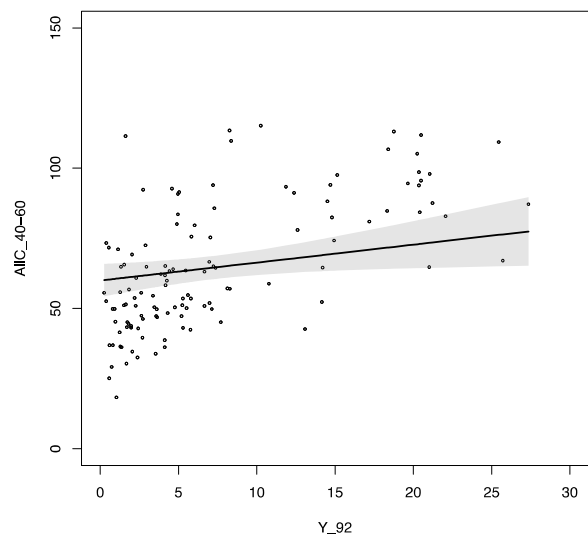


Figure A3. Marginal impact (and 95% confidence band) of p.c. income in the parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: 40-60 yrs.

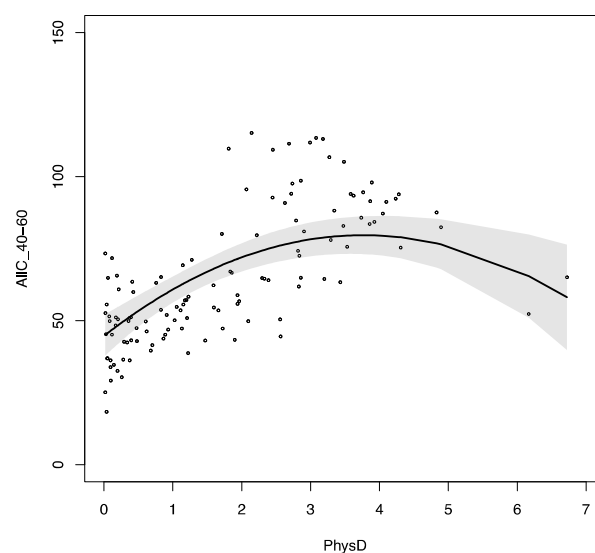


Figure A4. Marginal impact (and 95% confidence band) of physician density in the parametric regression of standardized incidence of all-sites cancers on p.c. income and physician density. Age class: 40-60 yrs.

724

Table A1. Descriptive statistics of the variables

	Min	Max	Median	Average	Stand Err.
AllC	61.8	374.1	182.2	196.3	78.9
AllC_40_60	18.3	115.2	60.4	64.7	22.7
Lung	0.2	55.5	15.6	17.8	13.2
Lung_40_60	0.1	22.8	4.2	5.1	4.2
Breast	5.4	118.5	45.1	51.4	27.4
Breast_40_60	2.7	57.7	23	25.3	12.8
Colorectum	1.2	48.8	13.1	18.1	13.4
Colorectum_40_60	0.3	16.6	3.9	5.1	3.2
Prostate	1.3	144.4	31.6	44.4	37
Prostate_40_60	0	67.5	4	7.5	9.2
Stomach	0.8	45.4	7.3	9.8	7.6
Stomach_40_60	0.3	15.4	2	2.8	2.4
Liver	1.1	89	5.2	7.7	9.5
Liver_40_60	0.2	29.7	1.4	2.5	3.4
Cervix	2.3	86.7	16.9	20.8	15.4
Cervix_40_60	1.2	49	8.4	9.9	7.7
Oesophagus	0	28	2.6	4.3	4.9
Oesophagus_40_60	0	10.3	0.8	1.3	1.5
Y_92	251	27352	4833	7266	6858
PhysD	0.02	6.72	1.59	1.80	1.50

725

Table A2: Correlation matrix of all-sites incidence rates and regressors

	AllC	AllC 40-60	Y_92	Y_97	Y_02	Y_07	Y_12	Phys
AllC	1							
AllC 40-60	0.93	1						
Y_92	0.78	0.64	1					
Y_97	0.78	0.63	0.99	1				
Y_02	0.80	0.65	0.98	0.99	1			
Y_07	0.80	0.65	0.96	0.98	0.99	1		
Y_12	0.80	0.66	0.95	0.97	0.98	0.99	1	
Phys	0.71	0.67	0.61	0.58	0.60	0.62	0.63	1

726

727

728

729 Table A3. Summary of the Spatial Error Model estimates for the 8 most common organ-sites cancer,  
 730 40\_60 age classes

1	2	3	4	5	6	7	8	9
<i>Organ Site</i>	<i>Rel. freq.</i>	<i>Typical of</i>	<i>Model</i>	<i>Role of Y_92</i>	<i>Role of PhysD</i>	<i>Y_92 coeff.</i>	<i>Y_92<sup>2</sup> coeff.</i>	<i>Adj RSq</i>
Lung	13.0%	H	Linear	none	+	n.s.		0.61
Breast	11.9%	H	Linear	+	+	0.80***		0.64
Colorectum	9.7%	H	EKC	\$	+	0.51***	-0.012**	0.67
Prostate	7.9%	H	Linear	+	n.s.	0.42***		0.47
Stomach	6.8%	M-L	Linear	-	+	-0.17***		0.35
Liver	5.6%	M-L	Linear	-	n.s.	-0.15**		0.11
Cervix uteri	3.7%	M-L	Linear	-	-	-0.26**		0.47
Oesophagus	3.2%	M-L	none	none	n.s.			

731 Significance p levels: \*\*<0.05, \*\*\*<0.01, ^>=0.11; n.s. = non significant

\$ calculated turning point = 21,450 \$ GDP 1992PPP

732

733

734

735

736

737

738

739

740

741

742Table A4: Regressions for the most frequent organ-site cancers

743				745			
<b>LUNG</b>				<b>STOMACH</b>			
ALL	Coeff.	Std. Error	p	ALL	Coeff.	Std. Error	p
Intercept	5.34	1.86	0.00	Intercept	5.37	1.53	0.00
Y_92	0.41	0.17	0.01	Y_92	-0.37	0.13	0.00
PhysD	7.40	1.73	0.00	PhysD	6.93	1.32	0.00
PhysD <sup>2</sup>	-0.71	0.30	0.02	PhysD <sup>2</sup>	-0.98	0.22	0.00
Spatial parameter=0.36*				Spatial parameter=0.46***			
AdjRsqr=0.61				AdjRsqr=0.36			
40-60	Coeff.	Std. Error	p	40-60	Coeff.	Std. Error	p
Intercept	2.07	0.72	0.00	Intercept	1.87	0.47	0.00
Y_92	-0.04	0.06	0.55	Y_92	-0.17	0.04	0.00
PhysD	2.62	0.62	0.00	PhysD	2.15	0.41	0.00
PhysD <sup>2</sup>	-0.26	0.10	0.01	PhysD <sup>2</sup>	-0.29	0.07	0.00
Spatial parameter=0.47***				Spatial parameter=0.45***			
AdjRsqr=0.61				; AdjRsqr=0.35			
<b>BREAST</b>				<b>LIVER</b>			
ALL	Coeff.	Std. Error	p	ALL	Coeff.	Std. Error	p
Intercept	22.67	3.87	0.00	Intercept	8.45	1.80	0.00
Y_92	1.89	0.32	0.00	Y_92	-0.36	0.18	0.04
PhysD	16.23	3.27	0.00	PhysD	0.99	0.77	0.20
PhysD <sup>2</sup>	-2.62	0.55	0.00	Spatial parameter=0.36**			
Spatial parameter=0.49***				AdjRsqr=0.08			
AdjRsqr=0.70				40-60	Coeff.	Std. Error	p
40-60	Coeff.	Std. Error	p	Intercept	2.89	0.65	0.00
Intercept	11.57	1.89	0.00	Y_92	-0.15	0.06	0.01
Y_92	0.80	0.16	0.00	PhysD	0.38	0.27	0.17
PhysD	8.93	1.65	0.00	Spatial parameter=0.38***			
PhysD <sup>2</sup>	-1.49	0.28	0.00	AdjRsqr=0.11			
Spatial parameter=0.45***							
AdjRsqr=0.64							
<b>COLORECTUM</b>				<b>CERVIX</b>			
ALL	Coeff.	Std. Error	p	ALL	Coeff.	Std. Error	p
Intercept	0.51	1.66	0.76	Intercept	31.11	2.99	0.00
Y_92	2.25	0.39	0.00	Y_92	-0.39	0.24	0.10
Y_92 <sup>2</sup>	-0.05	0.02	0.00	PhysD	-7.03	2.44	0.00
PhysD	5.96	1.42	0.00	PhysD <sup>2</sup>	0.90	0.41	0.03
PhysD <sup>2</sup>	-0.84	0.24	0.00	Spatial parameter=0.53***			
Spatial parameter=0.39***;				AdjRsqr=0.48			
AdjRsqr=0.76				40-60	Coeff.	Std. Error	p
40-60	Coeff.	Std. Error	p	Intercept	15.81	1.44	0.00
Intercept	0.82	0.46	0.07	Y_92	-0.26	0.12	0.03
Y_92	0.51	0.11	0.00	PhysD	-3.65	1.21	0.00
Y_92 <sup>2</sup>	-0.01	0.00	0.01	PhysD <sup>2</sup>	0.48	0.20	0.02
PhysD	1.77	0.40	0.00	Spatial parameter=0.49***			
PhysD <sup>2</sup>	-0.27	0.07	0.00	AdjRsqr=0.47			
Spatial parameter=0.36***							
AdjRsqr=0.67							
<b>PROSTATE</b>				<b>OESOPHAGUS</b>			
ALL	Coeff.	Std. Error	p	ALL	Coeff.	Std. Error	p
Intercept	21.45	6.53	0.00	Intercept	4.51	1.20	0.00
Y_92	2.43	0.42	0.00	Y_92	-0.21	0.24	0.38
PhysD	4.32	4.24	0.31	Y_92 <sup>2</sup>	0.003	0.01	0.77
PhysD <sup>2</sup>	-0.68	0.69	0.32	PhysD	0.50	0.37	0.18
Spatial parameter=0.70***				Spatial parameter=0.57***			
AdjRsqr=0.75				AdjRsqr=0.28			
40-60	Coeff.	Std. Error	p	40-60	Coeff.	Std. Error	p
Intercept	3.80	1.71	0.03	Intercept	1.49	0.37	0.00
Y_92	0.42	0.14	0.00	Y_92	-0.08	0.08	0.28
PhysD	0.32	0.59	0.59	Y_92 <sup>2</sup>	0.002	0.003	0.58
Spatial parameter=0.56***				PhysD	0.10	0.12	0.39
AdjRsqr=0.47				Spatial parameter=0.54***			

746

**748List of Countries and Regions**

749

**750COUNTRIES**

751

752Albania	783Ecuador	814Kenya	845Samoa
753Algeria	784Egypt	815Korea, Republic of	846Singapore
754Armenia	785El Salvador	816Kyrgyzstan	847South African Rep.
755Australia	786Ethiopia	817Lebanon	848Spain
756Austria	787Fiji	818Malawi	849Sri Lanka
757Azerbaijan	788Finland	819Malaysia	850Sudan
758Bahamas	789France (metrop.).	820Mali	851Suriname
759Bangladesh	790FYR Macedonia	821Malta	852Swaziland
760Barbados	791Gabon	822Mauritius	853Sweden
761Belarus	792Gambia	823Mexico	854Switzerland
762Belgium	793Georgia	824Moldova, rep. of	855Tajikistan
763Belize	794Germany	825Mongolia	856Tanzania
764Bhutan	795Ghana	826Morocco	857Thailand
765Bolivia	796Greece	827Mozambique	858Togo
766Botswana	797Guatemala	828Namibia	859Trin. and Tobago
767Brazil	798Guinea	829New Zealand	860Tunisia
768Bulgaria	799Guyana	830Nicaragua	861Turkey
769Burkina Faso	800Honduras	831Niger	862Turkmenistan
770Cameroon	801Hungary	832Nigeria	863Uganda
771Canada	802Iceland	833Netherlands, the	864Ukraine
772Chile	803India	834Norway	865United Kingdom
773China	804Indonesia	835Pakistan	866USA
774Colombia	805Iran, Islamic Rep. of	836Panama	867Uruguay
775Congo, Rep. of	806Iraq	837Papua New Guinea	868Uzbekistan
776Costa Rica	807Ireland	838Paraguay	869Vanuatu
777Cote d'Ivoire	808Israel	839Peru	870Venezuela
778Cuba	809Italy	840Philippines	871Vietnam
779Cyprus	810Jamaica	841Poland	872Yemen
780Czech Republic	811Japan	842Portugal	873Zambia
781Denmark	812Jordan	843Romania	
782Dominican Rep.	813Kazakhstan	844Russian Federation	

874

875REGIONS

876Australia/New Zealand, Caribbean, Central America, Eastern Africa, Eastern Asia, Eastern

877Europe, Mela/Micro/Polynesia, Middle Africa, Northern Africa, Northern America, Northern

878Europe, South America, South Central Asia, South Eastern Asia, Southern Africa, Southern

879Europe, Western Africa, Western Asia, Western Europe

880